

Moderate-to-Severe diarrhea 16S dataset

December 3, 2014

This data package contains the information used to run the analyses found in "Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition". Measurements are the number of reads annotated for a particular cluster within a given sample followed by filtering. Sequencing was performed on the 454 Flex platform. Data is stored as an `MRexperiment`-class object. The count matrix was generated using DNAClust (<http://dnacust.sourceforge.net/>). For more details please refer to the paper.

The help file `?msd16s` describes the example dataset.

1 The Data

We start by loading the library and the data.

```
> suppressMessages(library(metagenomeSeq))
> library(msd16s)
> data(msd16s)
```

This will load the `msd16s` object of class `MRexperiment`. As described in the `metagenomeSeq` vignette, `print` (or `show`) will display summary information.

```
> msd16s

MRexperiment (storageMode: environment)
assayData: 26044 features, 992 samples
  element names: counts
protocolData: none
phenoData
  sampleNames: 100259 100262 ... 602385 (992 total)
  varLabels: Type Country ... Dysentery (5 total)
  varMetadata: labelDescription
featureData
  featureNames: 54 94 ... 276421 (26044 total)
```

```

fvarLabels: OTU Taxonomy ... clusterCenter (10 total)
fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:

```

The data in `msd16s` is the substrate for the analysis described in "Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition". Included in the `MRexperiment` object are the counts, phenotype and feature information.

The phenotype information can be accessed with the `phenoData` and `pData` methods:

```
> phenoData(msd16s)
```

```

An object of class 'AnnotatedDataFrame'
sampleNames: 100259 100262 ... 602385 (992 total)
varLabels: Type Country ... Dysentery (5 total)
varMetadata: labelDescription

```

```
> head(pData(msd16s))
```

	Type	Country	Age	AgeFactor	Dysentery
100259	Case	Gambia	14	[12,18)	1
100262	Control	Gambia	24	[24,60)	0
100267	Case	Gambia	17	[12,18)	0
100274	Case	Gambia	36	[24,60)	0
100275	Case	Gambia	29	[24,60)	0
100277	Case	Gambia	29	[24,60)	0

The feature information including cluster representative sequence can be accessed with the `featureData` and `fData` methods:

```
> featureData(msd16s)
```

```

An object of class 'AnnotatedDataFrame'
featureNames: 54 94 ... 276421 (26044 total)
varLabels: OTU Taxonomy ... clusterCenter (10 total)
varMetadata: labelDescription

```

```
> head(fData(msd16s))
```

	OTU
54	54
94	94
113	113

```
117 117
145 145
202 202
```

```
54 ;cellular organisms;Bacteria;Firmicutes;Bacilli;Lactobacill
94 ;cellular organisms;Bacteria;Firmicutes;Bacilli;Lactobacill
113 ;cellular organisms;Bacteria;Firmicutes;Bacilli;Lactobacill
117 ;cellular organisms;Bacteria;Firmicutes;Bacilli;Lactobacill
145 ;cellular organisms;Bacteria;Bacteroidetes/Chlorobi group;Bacteroidetes;Bacteroidia
202 ;cellular organisms;Bacteria;Bacteroidetes/Chlorobi group;Bacteroidetes;Bacteroid
      superkingdom      phylum      class      order      family
54      Bacteria      Firmicutes      Bacilli      Lactobacillales      Lactobacillaceae
94      Bacteria      Firmicutes      Bacilli      Lactobacillales      Lactobacillaceae
113     Bacteria      Firmicutes      Bacilli      Lactobacillales      Lactobacillaceae
117     Bacteria      Firmicutes      Bacilli      Lactobacillales      Lactobacillaceae
145     Bacteria      Bacteroidetes      Bacteroidia      Bacteroidales      Prevotellaceae
202     Bacteria      Bacteroidetes      Bacteroidia      Bacteroidales      Bacteroidaceae
      genus
54      Lactobacillus      Lactobacillus sp. TSK G32-2
94      Lactobacillus      Lactobacillus sp. TSK G32-2
113     Lactobacillus      Lactobacillus sp. TSK G32-2
117     Lactobacillus      Lactobacillus sp. TSK G32-2
145     Prevotella        Prevotella sp. DJF_RP53
202     Bacteroides        Bacteroides fragilis
```

```
54      CATGCTGCCTCCCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCAACCTCTCAGTTCGGCTACGTATCATCAO
94      CATGCTGCCTCCCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCAACCTCTCA
113     CATGCTGCCTCCCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCAAC
117     CATGCTGCCTCCCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCAAC
145     CATGCTGCCTCCCGTAGGAGTTTGGACCGTGTCTCAGTTCCAATGTGGGGG
202     CATGCTGCCTCCCGTAGGAGTTTGGACCGTGTCTCAGT
```

The raw or normalized counts matrix can be accessed with the `MRcounts` function:

```
> head(MRcounts(msd16s[,1:10]))
```

```
      100259 100262 100267 100274 100275 100277 100291 100292 100293 100294
54      0      0      0      0      0      0      0      0      0      0
94      0      0      0      0      0      0      0      0      0      0
113     0      0      0      0      0      0      0      0      0      0
117     0      0      0      0      0      0      0      0      0      0
145     0      0      0      0      0      0      0      0      0      0
202     0      0      0      0      0      0      0      0      0      0
```

Using this class, the object can be easily subsetted, for example:

```
> msd16s_bangladesh = msd16s[,pData(msd16s)$Country == "Bangladesh"]  
> msd16s_bangladesh
```

```
MRexperiment (storageMode: environment)  
assayData: 26044 features, 206 samples  
  element names: counts  
protocolData: none  
phenoData  
  sampleNames: 600002 600005 ... 602385 (206 total)  
  varLabels: Type Country ... Dysentery (5 total)  
  varMetadata: labelDescription  
featureData  
  featureNames: 54 94 ... 276421 (26044 total)  
  fvarLabels: OTU Taxonomy ... clusterCenter (10 total)  
  fvarMetadata: labelDescription  
experimentData: use 'experimentData(object)'  
Annotation:
```